A nearly optimal discrete query quantum algorithm for evaluating NAND formulas

Andris Ambainis*

Abstract

We present an $O(\sqrt{N})$ discrete query quantum algorithm for evaluating balanced binary NAND formulas and an $O(N^{\frac{1}{2}+O(\frac{1}{\sqrt{\log N}})})$ discrete query quantum algorithm for evaluating arbitrary binary NAND formulas.

1 Introduction

One of two most famous quantum algorithms is Grover's search [15] which solves a generic problem of exhaustive search among N possibilities in $O(\sqrt{N})$ steps. This provides a quadratic speedup over the naive classical algorithm for a variety of search problems [3].

Grover's algorithm can be re-cast as computing OR of N bits $x_1, \ldots, x_N$, with $O(\sqrt{N})$ queries to a black box storing the values $x_1, \ldots, x_N$. Then, a natural generalization of this problem is computing the value of an AND-OR formula of $x_1, \ldots, x_N$.

This problem can be viewed as a black-box model for determining the winner in a 2-player game (such as chess) if both players play their optimal strategies. In this case, the game can be represented by a game tree consisting of possible positions. The leaves of a tree correspond to the possible end positions of the game. Each of them contains a variable x_i, with $x_i = 1$ if the first player wins and $x_i = 0$ otherwise. Internal nodes corresponding to positions which the first player makes the next move contain a value that is OR of the values of their children. (The first player wins if he has a move that leads to a position from which he can win.) Internal nodes for which the second player makes the next move contain a value that is AND of the values of their children. (The first player wins if he wins for any possible move of the second player.)

The question is: assuming we have no further information about the game beyond the position tree, how many of the variables x_i do we have to examine to determine whether the first player has a winning strategy? This problem has been studied in both classical [20, 19, 21] and quantum [2, 5, 8, 16] context.

Throughout this paper, we will assume that the formula is read-once (every leaf contains a different variable). There are two main cases that have been studied in the quantum case.

The first case is when the formula is of a constant depth d. If the formula is balanced (which is the most commonly studied case), then even levels contain ORs of $N^{1/d}$ variables and odd levels contain ANDs of $N^{1/d}$ variables. In this case, $\Theta(\sqrt{N})$ quantum queries are both sufficient [8, 16] and necessary [2]. Since any randomized algorithm requires $\Omega(N)$ queries in this case, we still achieve a quadratic speedup over the best classical algorithm. For arbitrary formulas of depth d, $O(\sqrt{N}\log^{d-1} N)$ queries suffice [5]. This is almost tight, as Barnum and Saks [7] have shown that $\Omega(\sqrt{N})$ queries are necessary to evaluate any AND-OR formula of any depth.

*Department of Combinatorics and Optimization and IQC, University of Waterloo, Canada, ambainis@math.uwaterloo.ca. Supported by NSERC, CIAR, ARO/DTO, MITACS and IQC.

The second case is when, instead of a constant depth, we have a constant fan-out. This case has been much harder and, until a few months ago, there has been no progress on it at all. If we restrict to binary AND-OR trees, the classical complexity of computing the value of a balanced binary AND-OR tree is $\Theta(N^{.754\cdots})$ [20, 19, 21] and there was no better quantum algorithm known.

In a breakthrough result, Farhi et al. [14] showed that the value of a balanced binary NAND tree can be computed in $O(\sqrt{N})$ quantum time in an unconventional continuous-time Hamiltonian query model of [13, 18]. (Because of De Morgan's laws, computing the value of an AND-OR tree is equivalent to computing the value of a NAND tree.) Using a standard reduction between continuous time and discrete time quantum computation [10], this yields an $O(N^{1/2+\epsilon})$ query quantum algorithm in the standard discrete time quantum query model, for any $\epsilon > 0$. (The big-O constant deteriorates, as the ϵ decreases.)

Soon after, Childs et al. [11] extended the result of [14] to computing the value of an arbitrary binary NAND tree of depth in time $O(\sqrt{Nd})$ in the continuous-time Hamiltonian query model and with $O(N^{1/2+\epsilon})$ queries in the discrete-time query model.

In this paper, we improve over [14] and [11] by giving a better discrete time quantum query algorithms for both balanced and general NAND trees. Namely, we give

1. An $O(\sqrt{N})$ query quantum algorithm for evaluating balanced binary NAND formulas, which is optimal up to a constant factor.
2. An $O(\sqrt{Nd})$ query quantum algorithm for evaluating arbitrary binary NAND formulas of depth d.
3. An $O(N^{\frac{1}{2}+O(\frac{1}{\sqrt{\log N}})})$ query quantum algorithm for evaluating arbitrary binary NAND formulas of any depth.

All of our algorithms are designed directly in the discrete quantum query model and do not incur the overhead from converting from continuous to discrete time.

Besides better running time, our algorithms provide a new perspective for understanding the quantum algorithms for this problem. When the breakthrough algorithm of [14] appeared, its ideas seemed to be very different from anything known before. Our new algorithm and its analysis show intricate connections to the previous work on quantum search.

Although its technical details are complex, the main intuition is the same as in Grover's search [15] and its "two reflections" analysis [1] which views the Grover's algorithm as a sequence of reflections in two-dimensional space against two different axes. The idea of "two reflections" has come up in quantum algorithms over and over. For example, the element distinctness algorithm of [4], designed by different methods, was re-cast in the form of two reflections by Szegedy [22]. In this paper, we show that the NAND-tree algorithms can be viewed as another instance of "two-reflections", with the reflections designed, using the structure of the NAND tree.

2 Preliminaries

2.1 Quantum query model

We work in the standard discrete time quantum query model [3, 9]. In this model, the input bits can be accessed by queries to an oracle X and the complexity of f is the number of queries needed to compute f. A quantum computation with T queries is just a sequence of unitary transformations

$$V_0 \to O \to V_1 \to O \to \ldots \to V_{T-1} \to O \to V_T.$$

The V_j's can be arbitrary unitary transformations that do not depend on the input bits $x_1, \ldots, x_N$. The O's are query (oracle) transformations which depend on $x_1, \ldots, x_N$. To define

O, we represent basis states as $|i,z\rangle$ where $i \in \{0,1,\ldots,N\}$. The query transformation O_x (where $x=(x_1,\ldots,x_N)$) maps $|0,z\rangle$ to $|0,z\rangle$ and $|i,z\rangle$ to $(-1)^{x_i}|i,z\rangle$ for $i \in \{1,...,N\}$ (i.e., we change phase depending on x_i, unless $i=0$ in which case we do nothing).

The computation starts with a state $|0\rangle$. Then, we apply V_0, O_x, ..., O_x, V_T and measure the final state. The result of the computation is the rightmost bit of the result of the measurement. A quantum algorithm computes a function $f(x_1,\ldots,x_N)$ if, for any $x_1,\ldots,x_N \in \{0,1\}$, the probability that the result of the measurement is equal to $f(x_1,\ldots,x_N)$ is at least $2/3$.

We will describe our algorithm in a high level language but it can be translated into a sequence of transformations of this form.

2.2 Phase estimation

In our algorithm, we use phase estimation [12]. Assume that we are given a black box performing a unitary transformation U and a state $|\psi\rangle$ which is an eigenstate of U: $U|\psi\rangle = e^{i\theta}|\psi\rangle$. Our goal is to obtain an estimate $\tilde{\theta}$ such that $|\tilde{\theta}-\theta| < \delta$ with probability at least $1-\epsilon$. The algorithm for phase estimation by [12] solves this problem by invoking U $O(\frac{1}{\delta\epsilon})$ times.

If the input to this algorithm is a state $|\psi\rangle$ that is a linear combination of different eigenstates: $|\psi\rangle = \sum_j \alpha_j|\psi_j\rangle$ with $U_i|\psi_j\rangle = e^{i\theta_j}|\psi_j\rangle$, then the algorithm works as if the input was a probabilistic combination of $|\psi_j\rangle$ with probabilities $|\alpha_j|^2$.

3 Summary of results and methods

3.1 Results

Let T be a read-once binary NAND formula involving variables x_1, x_2, ..., x_N. We can represent T by a tree that have variables $x_1,\ldots,x_N$ at the leaves and NAND gates at the internal nodes. Let d be the depth of T. We have

Theorem 1 *1. If T is the complete binary tree, then $T(x_1,\ldots,x_N)$ can be computed with $O(\sqrt{N})$ quantum queries.*

2. For any binary tree T, $T(x_1,\ldots,x_N)$ can be computed with $O(\sqrt{dN})$ quantum queries.

We refer to the first part of the theorem as the *balanced case* and to the second part as the *general case.*

Bshouty et al. [6] have shown

Theorem 2 *[6] For any NAND formula T of size S, there exists a NAND formula T' of size $S' = O(S^{1+O(\frac{1}{\sqrt{\log S}})})$ and depth $d = O(S^{O(\frac{1}{\sqrt{\log S}})})$ such that $T' = T$.*

This theorem follows by substituting $k = 2^{\frac{1}{\sqrt{\log S}}}$ into Theorem 6 of [6]. By combining Theorems 1 and 2, we have

Corollary 1 *For any T, $T(x_1,\ldots,x_N)$ can be computed with $O(N^{\frac{1}{2}+O(\frac{1}{\sqrt{\log N}})})$ quantum queries.*

If the formula T is not read once, the number of variables N is replaced by the size of the formula S. This gives us

Corollary 2 *If $T(x_1,\ldots,x_N)$ is computable by a NAND formula of size S, T is computable by a quantum query algorithm with $O(S^{\frac{1}{2}+O(\frac{1}{\sqrt{\log S}})})$ queries.*

The link between quantum query complexity and formula size was first noticed by Laplante et al. [17] who observed that, whenever quantum adversary lower bound method of [2] gives a lower bound of $\Omega(M)$ for quantum query algorithms, it also gives a lower bound of $\Omega(M^2)$ for formula size. Based on that, they conjectured that any Boolean function with formula size M^2 has a quantum query algorithm with $O(\sqrt{M})$ queries. The results in [11] and this paper show that it is indeed possible to transform an arbitrary NAND formula into a quantum query algorithm, with almost a quadratic relation between formula size and the number of queries.

3.2 The algorithm

Our algorithm is the same for both parts of Theorem 3. Without the loss of generality, assume that all leafs are at an even distance from the root. (If there is a leaf l at an odd depth, create two new vertices v_1, v_2 and connect them to l, making l an internal node. v_1, v_2 are now leaves at an even depth. If x_i is the variable that used to be at the leaf l, replace it by two new variables at leaves v_1, v_2 and make both of those equal to $NOTx_i$. Then, the NAND of those two variables at the vertex l will evaluate to x_i.)

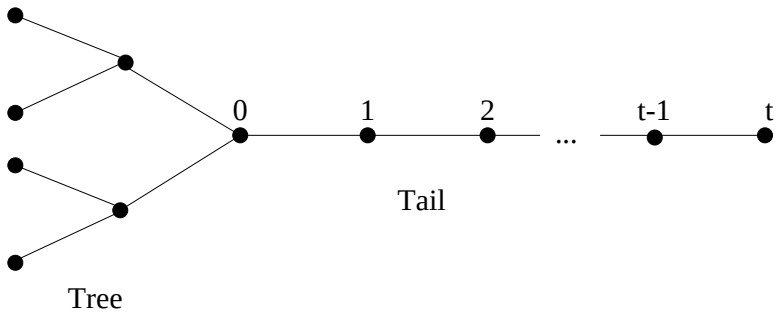

Figure 1: A tree T, augmented by a tail.

We augment the tree T by a "tail" of an even length t where $t = 2\lceil\sqrt{N}\rceil$ in the balanced case and $t = 2\lceil\sqrt{Nd}\rceil$ in the general case. The tail is a path that starts at the root of T and then goes through t newly created vertices. T' denotes the tree T, augmented by the tail (see Figure 1).

Our state space $\mathcal{H}$ will be spanned by basis states $|v\rangle$ corresponding to vertices of T'. We use $|i\rangle$ (for $i = 0, \ldots, t$) to denote the basis state corresponding to the i^{th} vertex in the tail of T'. $|0\rangle$ corresponds to the root of T.

We define a Hermitian matrix H as follows:

1. If pc is an edge in T from a parent p to a child c, then $H_{pc} = H_{cp} = \sqrt[4]{\frac{m_p}{2m_c}}$ if p is at an odd level and $H_{pc} = H_{cp} = \sqrt[4]{\frac{2m_c}{m_p}}$ if p is at an even level. (When T is the complete balanced tree, this becomes $H_{cp} = H_{pc} = 1$.)
2. If uv is an edge in the tail, then $H_{uv} = H_{vu} = 1$.
3. If uv is not an edge, then $H_{uv} = H_{vu} = 0$.

Then, H is a Hermitian operator acting on $\mathcal{H}$. Let $S_{H,0}$ be the 0-eigenspace of H, $S_{H,1} = (S_{H,0})^{\perp}$ and let U_1 be defined by $U_1|\psi\rangle = |\psi\rangle$ for $|\psi\rangle \in S_{H,0}$ and $U_1|\psi\rangle = -|\psi\rangle$ for $|\psi\rangle \in S_{H,1}$.

Let U_2 be defined by $U_2|\psi\rangle = -|\psi\rangle$ if $|\psi\rangle$ belongs to the subspace $S_{x,1}$ spanned by basis states $|v\rangle$ that correspond to leaves containing variables $x_i = 1$ and $U_2|\psi\rangle = |\psi\rangle$ if $|\psi\rangle$ belongs to the subspace $S_{x,0}$ spanned by all other basis states $|v\rangle$.

U_2 can be implemented with one query O. (It is essentially the query transformation O_x, with basis states labelled in a different way.) U_1 is independent of $x_1, \ldots, x_N$ and can be implemented without using the query transformation O.

Let $|\psi_{start}\rangle = \sum_{i=0}^{t/2} |2i\rangle$. (This is the starting state for the continuous time algorithm of Childs et al. [11].) Let $|\psi'_{start}\rangle = P_{S_{H,0}}|\psi_{start}\rangle$ and let $|\psi''_{start}\rangle = \frac{|\psi'_{start}\rangle}{\||\psi'_{start}\|}$.

Theorem 3 *1. If T evaluates to 0, there is a state $|\psi_0\rangle$ such that $U_2U_1|\psi_0\rangle = |\psi_0\rangle$ and $|\langle\psi_0|\psi''_{start}\rangle|^2 \geq c$ for some constant $c > 0$.*

2. If T evaluates to 1, then, for any eigenstate $|\psi_0\rangle$ of U_2U_1 which is not orthogonal to $|\psi''_{start}\rangle$, the corresponding eigenvalue of U_2U_1 is $e^{i\theta}$, with $\theta = \Omega(\frac{1}{\sqrt{Nd}})$ for any T and $\theta = \Omega(\frac{1}{\sqrt{N}})$ when T is the complete balanced tree.

We can distinguish the two cases by running the eigenvalue estimation for U_2U_1, with $|\psi''_{start}\rangle$ as the starting state, precision $\delta = \frac{\theta_{min}}{2}$ where θ_{min} is the lower bound on θ from the second part of Theorem 3 ($\theta_{min} = \Theta(\frac{1}{\sqrt{Nd}})$ or $\theta_{min} = \Theta(\frac{1}{\sqrt{N}})$) and error probability $\epsilon \leq \frac{c}{3}$. In the first case, with probability $|\langle\psi_0|\psi''_{start}\rangle|^2 \geq c$, we get the same answer as if the input to eigenvalue estimation was $|\psi_0\rangle$. Since the correct eigenvalue is 0, this means that we get an answer $\tilde{\theta} < \frac{\theta_{min}}{2}$ with probability at least $(1-\epsilon)c$.

In the second case, if we write out $|\psi''_{start}\rangle$ as a linear combination of eigenvectors of U_2U_1, all of those eigenvectors have eigenvalues that are $e^{i\theta}$, $\theta > \theta_{min}$. Therefore, the probability of the eigenvalue estimation outputting an estimate $\tilde{\theta} < \theta_{min} - \delta = \frac{\theta_{min}}{2}$ is at most ϵ.

By our choice of ϵ, we have $(1-\epsilon)c > \epsilon$. We can distinguish the two cases with arbitrarily high probability, by repeating the eigenvalue estimation C times, for a sufficiently large constant C.

3.3 Proof overview

The first part of Theorem 3 is proven by constructing the state $|\psi_0\rangle$. For the second part, we show that the entire state-space $\mathcal{H}$ can be expressed as a direct sum of one-dimensional and two-dimensional subspaces, with each subspace being mapped to itself by U_1 and U_2. Each one-dimensional subspace consists of all multiples of some state $|\psi\rangle$, with $U_1|\psi\rangle$ and $U_2|\psi\rangle$ being either $|\psi\rangle$ or $-|\psi\rangle$. Therefore, we either have $U_2U_1|\psi\rangle = |\psi\rangle$ or $U_2U_1|\psi\rangle = -|\psi\rangle$. We show that, if $U_2U_1|\psi\rangle = |\psi\rangle$, then $|\psi\rangle$ is orthogonal to the starting state $|\psi''_{start}\rangle$ and, therefore, has no effect on the algorithm.

For two-dimensional subspaces, we show that each of them has an orthonormal basis $|\psi_{11}\rangle$, $|\psi_{12}\rangle$ such that $U_1|\psi_{11}\rangle = |\psi_{11}\rangle$ and $U_1|\psi_{12}\rangle = -|\psi_{12}\rangle$ and another orthonormal basis $|\psi_{21}\rangle$, $|\psi_{22}\rangle$ such that $U_2|\psi_{21}\rangle = |\psi_{21}\rangle$ and $U_2|\psi_{22}\rangle = -|\psi_{22}\rangle$. Then, on this two-dimensional subspace, U_2U_1 is a product of two reflections, one w.r.t. $|\psi_{11}\rangle$ and one w.r.t. $|\psi_{21}\rangle$. As in "two reflections" analysis [1] of Grover's search, a product of two reflections in a two-dimensional plane is a rotation of plane by 2β, where β is the angle between $|\psi_{11}\rangle$ and $|\psi_{21}\rangle$. A rotation of the plane by 2β has eigenvalues $e^{\pm i\beta}$. Therefore, we need to lower-bound β.

Since $|\psi_{21}\rangle$ and $|\psi_{22}\rangle$ are orthogonal, the angle between $|\psi_{11}\rangle$ and $|\psi_{22}\rangle$ is $\frac{\pi}{2} - \beta$. Therefore, $|\langle\psi_{22}|\psi_{11}\rangle| = \sin\beta$. Since $|\psi_{22}\rangle$ belongs to $S_{x,1}$ and $|\psi_{11}\rangle$ belongs to $S_{H,0}$, we have

$$|\langle\psi_{22}|\psi_{11}\rangle| = \|P_{S_{x,1}}|\psi_{11}\rangle\| \geq \min_{|\psi\rangle \in S_{H,0}} \|P_{S_{x,1}}|\psi\rangle\|.$$

Therefore, to lower-bound $\sin\beta$ and β, it suffices to lower-bound the minimum of $\|P_{S_{x,1}}|\psi\rangle\|$ for $|\psi\rangle \in S_{H,0}$. We do that by an induction over the depth of the tree.

4 Notation

In this section, we summarize the main notation used in this paper:

Trees. T is the tree which we are evaluating. T' is the tree T with the tail attached to it. T_v is the subtree of T rooted at v. We also use T (or T_v) to denote the Boolean function defined by evaluating the NAND tree T (or T_v).

m_v and d_v denote the number of leaves and the depth of T_v. r denotes the root of T. Thus, $T_r = T$.

Matrices. H is the weighted version of the adjacency matrix of T', defined in section 3.2. H_v is the restriction of H to rows and columns in T_v.

Subspaces. $S_{H,0}$ is the eigenspace of H with the eigenvalue 0. $S_{H,1}$ is the orthogonal complement of $S_{H,0}$: $S_{H,1} = (S_{H,0})^{\perp}$. $S_{v,0}$ denotes the 0-eigenspace of H_v. $S'_{H,0}$ and $S'_{v,0}$ are subspaces of $S_{H,0}$ and $S_{v,0}$, defined in section 7.

$S_{x,1}$ is the subspace spanned by $|v\rangle$, for all leaves v that correspond to a variable $x_i = 1$. $S_{x,0}$ is the subspace spanned by all other $|v\rangle$ (for v that are either leaves corresponding to $x_i = 0$ or non-leaves).

Unitary transformations U_1 is defined by $U_1|\psi\rangle = |\psi\rangle$ for $|\psi\rangle \in S_{H,0}$ and $U_1|\psi\rangle = -|\psi\rangle$ for $|\psi\rangle \in S_{H,1}$. U_2 is the query transformation. It can be equivalently described by defining $U_2|\psi\rangle = |\psi\rangle$ for $|\psi\rangle \in S_{x,0}$ and $U_2|\psi\rangle = -|\psi\rangle$ for $|\psi\rangle \in S_{x,1}$.

5 Structure of minimal certificates of T_v

Let C be a minimal certificate of $T_v = 0$. We would like to determine the structure of C. Let z_1 and z_2 be the two children of v and y_1, y_2 (y_3, y_4) be the children of z_1 (z_2, respectively). For $T_v = 0$, we need to have $T_{z_1} = T_{z_2} = 1$ which is equivalent to at least one of T_{y_1} and T_{y_2} and at least one of T_{y_3} and T_{y_4} evaluates to 0. Thus, a minimal 0-certificate for T_v consists of a minimal 0-certificate for one of $T_{y_1} = 0$ and $T_{y_2} = 0$ and a minimal 0-certificate for one of $T_{y_3} = 0$ and $T_{y_4} = 0$. Each of those 0-certificates can be decomposed in a similar way.

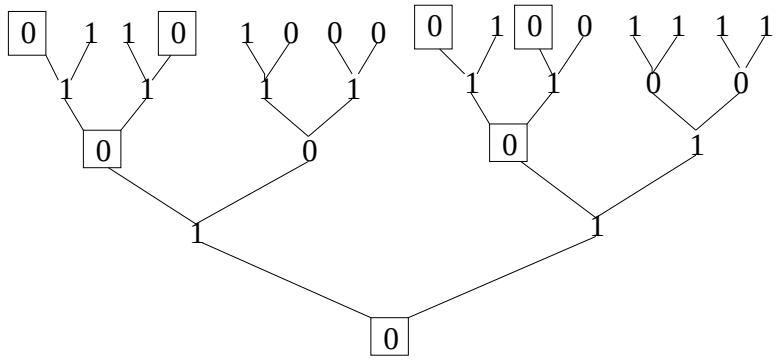

Figure 2: An extended certificate for $T_v = 0$.

We now define an extended minimal certificate for $T_v = 0$ to consist of v, an extended minimal certificate for one of $T_{y_1} = 0$ and $T_{y_2} = 0$ and an extended minimal 0-certificate for one of $T_{y_3} = 0$ and $T_{y_4} = 0$. Intuitively, an extended minimal certificate is a minimal certificate, augmented by non-leaf vertices that must evaluate to 0 on this certificate. We can show

Lemma 1 *Let C be an extended minimal certificate. If a non-leaf vertex w belongs to C, z_1 and z_2 are the two children of w and y_1, y_2 (y_3, y_4) are the children of z_1 (z_2, respectively), then exactly one of y_1, y_2 and exactly one of y_3, y_4 belongs to C.*

Proof: In appendix A. ∎

We show an example of an extended certificate for $T_v = 0$ in figure 2. The vertices that belong to the extended certificate are shown by squares.

For each extended minimal certificate C of $T_v = 0$, we can define a state $|\psi_C\rangle$ that has non-zero amplitudes only in the vertices of C, in a following way:

1. Decompose C as $C = C_{y_i} \cup C_{y_j} \cup \{v\}$, where C_{y_i} is an extended certificate for $T_{y_i} = 0$, $i \in \{1, 2\}$ and C_{y_j} is an extended certificate for $T_{y_j} = 0$, $j \in \{3, 4\}$.
2. Construct $|\psi_{C_{y_i}}\rangle$ and $|\psi_{C_{y_j}}\rangle$ inductively and define

$$|\psi_C\rangle = |v\rangle - \frac{\sqrt[4]{m_v}}{\sqrt[4]{4m_{y_i}}}|\psi_{C_{y_i}}\rangle - \frac{\sqrt[4]{m_v}}{\sqrt[4]{4m_{y_j}}}|\psi_{C_{y_j}}\rangle. \quad (1)$$

Lemma 2 *Let C be an extended minimal certificate for $T_v = 0$. Then,*

$$H_v|\psi_C\rangle = 0.$$

Proof: In appendix A. ∎

Lemma 3 *(a) If T is balanced, $\|\psi_{C_v}\|^2 \leq 2\sqrt{m_v} - 1$.*
(b) For any T, $\|\psi_{C_v}\|^2 \leq 2\sqrt{m_v d_v}$.

Proof: In appendix A. ∎

6 Proof of Theorem 1: $T = 0$ case

Let C be an extended minimal 0-certificate of T_r where r is the root of the tree and let $|\psi_C\rangle$ be the corresponding state (defined so that the amplitude α_r of the root is 1). We define $|\psi_0\rangle = |\psi_C\rangle + \sum_{i=1}^{t/2}(-1)^i|2i\rangle$. Let $|\psi_0'\rangle$ be the corresponding normalized state: $|\psi_0'\rangle = \frac{|\psi_0\rangle}{\||\psi_0\|}$.

We claim that $U_2U_1|\psi_0\rangle = |\psi_0\rangle$. This follows from $U_2|\psi\rangle = |\psi\rangle$ (which is true, because $|\psi_C\rangle$ and $|\psi_0\rangle$ are only non-zero on the vertices that belong to the extended certificate C and $x_i = 0$ for all variables x_i at the leaves that belong to a certificate C) and $U_1|\psi_0\rangle = |\psi_0\rangle$ (which follows from the next lemma).

Lemma 4

$$H|\psi_0\rangle = 0.$$

Proof: It suffices to show that, for every u, the amplitude of u in $H|\psi_0\rangle$ is 0. For vertices in the tree T, their amplitudes in $H|\psi_0\rangle$ are the same as their amplitudes in $H_r|\psi_C\rangle$ and, by Lemma 2, $H_r|\psi_C\rangle = 0$.

For vertices j in the tail, the amplitude of j in $H|\psi_0\rangle$ is the sum of the amplitudes of its two neighbors of $j-1$ and $j+1$ in $|\psi_0\rangle$. If $j = 2i$ is even, then both $2i-1$ and $2i+1$ have amplitudes 0 and their sum is 0. If $j = 2i+1$ is odd, then one of $2i$ and $2i+2$ has amplitude 1 and the other has amplitude -1, resulting in the sum of amplitudes being 0. ∎

To complete the proof of the first part of Theorem 3, we show

Lemma 5 *If $t > \sqrt{N}$ (for the balanced case) or $t > \sqrt{Nd}$ (for the unbalanced case), then*

$$\langle\psi_0'|\psi_{start}''\rangle \geq \frac{1}{\sqrt{5}}.$$

Proof: We show the proof for the balanced case. (For the unbalanced case, just replace N by Nd everywhere.)

Since $|\psi'_0\rangle \in S_{H,0}$ (by Lemma 4), we have

$$\langle\psi'_0|\psi_{start}\rangle = \langle\psi'_0|P_{S_{H,0}}|\psi_{start}\rangle = \|\psi'_{start}\|\langle\psi'_0|\psi''_{start}\rangle \leq \langle\psi'_0|\psi''_{start}\rangle.$$

Therefore, it suffices to prove $\langle\psi'_0|\psi_{start}\rangle \geq \frac{1}{\sqrt{5}}$. We have

$$\langle\psi'_0|\psi_{start}\rangle = \frac{\langle\psi_0|\psi_{start}\rangle}{\|\psi_0\|}.$$

Since each of the basis states $|2j\rangle$ has amplitude 1 in $|\psi_0\rangle$ and amplitude $\frac{1}{\sqrt{\frac{t}{2}+1}}$ in $|\psi_{start}\rangle$, we have $\langle\psi_0|\psi_{start}\rangle = \sqrt{\frac{t}{2}+1}$. We also have

$$\|\psi_0\|^2 = \|\psi_C\|^2 + \frac{t}{2} \leq 2\sqrt{N} + \frac{t}{2} \leq 2.5t.$$

Therefore,

$$\frac{\langle\psi_0|\psi_{start}\rangle}{\|\psi_0\|} \geq \frac{\sqrt{\frac{t}{2}+1}}{\sqrt{2.5t}} = \frac{1}{\sqrt{5}}.$$

∎

$\langle\psi'_0|\psi''_{start}\rangle$ can be increased to $1-\epsilon$ by taking $t \geq C\sqrt{N}$ for sufficiently large constant C.

7 Proof of Theorem 1: $T = 1$ case

7.1 Overview

We first describe a subset of the 1-eigenstates $|\psi\rangle$ of U_2U_1. Let v be a vertex of an odd depth $2j+1$ and let v_1 and v_2 be the children of v. Assume that $T_{v_1} = T_{v_2} = 0$ and let C_1, C_2 be the extended minimal certificates for $T_{v_1} = 0$ and $T_{v_2} = 0$. Define $|\psi_{C_1,C_2}\rangle = \sqrt[4]{m_{v_1}}|\psi_{C_1}\rangle - \sqrt[4]{m_{v_2}}|\psi_{C_2}\rangle$.

Lemma 6 $U_2|\psi_{C_1,C_2}\rangle = U_1|\psi_{C_1,C_2}\rangle = |\psi_{C_1,C_2}\rangle$.

Proof: In appendix A. ∎

We define $S'_{H,0}$ to be the subspace spanned by all $|\psi_{C_1,C_2}\rangle$, for all possible choices of v, C_1 and C_2. Observe that $S'_{H,0}$ is orthogonal to the starting state $|\psi_{start}\rangle$ (since the state $|\psi_{start}\rangle$ only has non-zero amplitude on the root and in the tail and any of the states $|\psi_{C_1,C_2}\rangle$ always has zero amplitudes there). Since $S'_{H,0} \subseteq S_{H,0}$, $S'_{H,0}$ is also orthogonal to $|\psi'_{start}\rangle = P_{S_{H,0}}|\psi_{start}\rangle$ and $|\psi''_{start}\rangle = \frac{|\psi'_{start}\rangle}{\|\psi'_{start}\|}$.

Theorem 3 follows from the following two lemmas:

Lemma 7 *Let S be a Hilbert space, let S_1, S_2 be two subspaces of S and let U_i ($i \in \{1,2\}$) be the unitary transformation on S defined by $U_i|\psi\rangle = -|\psi\rangle$ for $|\psi\rangle \in S_i$ and $U_i|\psi\rangle = |\psi\rangle$ for $|\psi\rangle \in (S_i)^\perp$. Assume that $S_1 \cap S_2 = \{\vec{0}\}$ and, for any $|\psi\rangle \in S_1$, we have*

$$\|P_{(S_2)^\perp}|\psi\rangle\|^2 \geq \epsilon.$$

Then, all eigenvalues of U_2U_1 are of the form $e^{i\theta}$ with $\theta \in [\sqrt{\epsilon}, 2\pi - \sqrt{\epsilon}]$.

Lemma 8 *For any state $|\psi\rangle \in S_{H,0} \cap (S'_{H,0})^\perp$, $\|P_{S_{x,1}}|\psi\rangle\|^2 \geq c\|\psi\|^2$ for a constant c, where $c = \Omega(\frac{1}{N})$ in the balanced case and $c = \Omega(\frac{1}{Nd})$ in the general case.*

Given the two lemmas, the proof of Theorem is completed as follows. Define $S = (S'_{H,0})^{\perp} \cap (S_{H,1} \cap S_{x,1})^{\perp}$. For both $S'_{H,0}$ and $S_{H,1} \cap S_{x,1}$, U_1 and U_2 map those subspaces to themselves. Since U_1 and U_2 are unitary, this means that U_1 and U_2 map S to itself, as well. Combining Lemma 7 and 8 implies that all eigenvalues of U_2U_1 on S are $e^{i\theta}$ with $\theta \in [\delta, 2\pi - \delta]$, with $\delta = \Omega(\frac{1}{\sqrt{N}})$ in the balanced case and $\delta = \Omega(\frac{1}{\sqrt{Nd}})$ in the general case.

Since $S'_{H,0}$ and $S_{H,1}$ are both orthogonal to $|\psi''_{start}\rangle$, the state $|\psi''_{start}\rangle$ belongs to S. Therefore, all eigenvectors of U_2U_1 which are not orthogonal to $|\psi''_{start}\rangle$ must lie in S and, therefore, have eigenvalues $e^{i\theta}$ of the required form. This completes the proof of the theorem.

Lemma 7 is fairly similar to previous work. We give its proof in Appendix B. In the next section, we give the proof of Lemma 8, postponing some of the technical details till the appendices.

7.2 Proof of Lemma 8

This lemma is equivalent to the next one.

Lemma 9 *For any state $|\psi\rangle \in S_{H,0}$, there exists a state $|\psi'\rangle \in S'_{H,0}$ such that*

$$\|P_{S_{x,1}}(|\psi\rangle + |\psi'\rangle)\|^2 \geq c\||\psi\rangle + |\psi'\rangle\|^2$$

for a constant c, where $c = \Omega(\frac{1}{N})$ in the balanced case and $c = \Omega(\frac{1}{Nd})$ in the general case.

Proof: [of Lemma 8] By construction of $S'_{H,0}$, any state in $S'_{H,0}$ is orthogonal to $S_{x,1}$. Let $|\psi\rangle \in S_{H_0} \cap (S'_{H,0})^{\perp}$. We use Lemma 9 to obtain $|\psi'\rangle \in S'_{H,0}$. Then, we have

$$\|P_{S_{x,1}}|\psi\rangle\|^2 = \|P_{S_{x,1}}(\psi + \psi')\|^2 \geq c\|\psi + \psi'\|^2 = c\|\psi\|^2 + c\|\psi'\|^2 \geq c\|\psi\|^2,$$

with the first equality following from $|\psi'\rangle$ being orthogonal to $S_{x,1}$, the second inequality from lemma 9 and the the third equality following from $|\psi\rangle$ being orthogonal to $|\psi'\rangle$ (which is true because $|\psi'\rangle \in S'_{H,0}$ but $|\psi\rangle \in (S'_{H,0})^{\perp}$). ∎

Proof: [of Lemma 9] We recall that $S_{v,0}$ is the 0-eigenspace of H_v. We define $S'_{v,0}$ to be the subspace spanned by all $|\psi_{C_1,C_2}\rangle$ which only have non-zero amplitudes for $|u\rangle$, $u \in T_v$. Then, we have $S'_{v,0} \subseteq S_{v,0}$.

For a state $|\psi\rangle$, we define

$$L_v(\psi) = \frac{\|\psi\|^2 - K\|P_{S_{x,1}}\psi\|^2}{|\alpha_v|^2}$$

where α_v is the amplitude of v of T_k in $|\psi\rangle$ and K will be defined later.

The next lemma is slightly different for balanced NAND trees and general NAND trees. Below is the variant for balanced trees. The counterpart for general trees is described in appendix D.

Lemma 10 *Let $K \geq 20N$ and let v be a vertex at depth $2k$. For any state $|\psi\rangle \in S_{v,0}$, there exists a state $|\psi'\rangle \in S'_{v,0}$ such that, for $|\psi''\rangle = |\psi\rangle + |\psi'\rangle$, we have*

1. *If $T_k(x_1, \ldots, x_N) = 0$, then*

$$L_v(\psi) \leq a_v, \quad a_v = \left(1 + 2\frac{2^{2k}}{K}\right)(2^{k+1} - 1).$$

2. *If $T_k(x_1, \ldots, x_N) = 1$, then*

$$L_v(\psi) \leq -b_v, \quad b_v = \left(1 - 2\frac{2^{2k}}{K}\right)\frac{K}{2^k}.$$

Given Lemma 10, the proof of Lemma 9 can be completed as follows. Let $|\psi\rangle$ be a 0-eigenstate of H. Then, we can decompose $|\psi\rangle = |\psi_{tree}\rangle + |\psi_{tail}\rangle$, with $|\psi_{tree}\rangle$ being a superposition over vertices in T (including the root) and $|\psi_{tail}\rangle$ being a superposition over vertices in the tail.

Let α_r be the amplitude of the root in $|\psi\rangle$. In both balanced and general case, we have

Lemma 11 *Let $|\psi\rangle = \sum_u \alpha_u |u\rangle$ be a 0-eigenstate of H (or H_v, for some v). Then, any vertex u of an odd depth (or any u in the tail at an odd distance from the root) must have $\alpha_u = 0$.*

Proof: In appendix A. ∎

By this lemma, $H|\psi\rangle = 0$ implies that the amplitudes of the vertices in the tail at an odd distance from the root are 0. Also, to achieve $H|\psi\rangle = 0$, the amplitudes at an even distance from the root must be $\pm\alpha_r$. Therefore, $\|\psi_{tail}\|^2 = \frac{t}{2}|\alpha_r|^2$.

We have $H_r|\psi_{tree}\rangle = 0$. Therefore, we can apply Lemma 10 with $v = r$ and $|\psi_{tree}\rangle$ instead of $|\psi\rangle$, obtaining a state $|\psi'\rangle \in S'_{r,0} \subseteq S'_{H,0}$. We let $|\psi''_{tree}\rangle = |\psi_{tree}\rangle + |\psi'\rangle$ and $|\psi''_{all}\rangle = |\psi\rangle + |\psi'\rangle = |\psi''_{tree}\rangle + |\psi_{tail}\rangle$. Then,

$$\|\psi''_{all}\|^2 = \|\psi''_{tree}\|^2 + \frac{t}{2}|\alpha_r|^2.$$

We have $P_{S_{x,1}}|\psi''_{all}\rangle = P_{S_{x,1}}|\psi''_{tree}\rangle$. From Lemma 10,

$$\|\psi''_{tree}\|^2 \leq K\|P_{S_{x,1}}|\psi''_{tree}\rangle\| - b_r|\alpha_r|^2.$$

This means that

$$\|\psi''_{all}\|^2 + \left(b_r - \frac{t}{2}\right)|\alpha_r|^2 \leq K\|P_{S_{x,1}}|\psi''_{all}\rangle\|.$$

If $t = 2\lceil\sqrt{N}\rceil$, we have $b_r \geq \frac{t}{2}$ (this can be verified by substituting the definition of b_r). Therefore, $\|P_{S_{x,1}}|\psi''_{all}\rangle\| \geq \frac{\|\psi''_{all}\|^2}{K}$.

The balanced case is based on similar ideas but has some minor changes in the expressions that appear in Lemma 10 (with the main change being $K \geq 20N$ replaced by $K \geq 30Nd$, which results in a bound of $c = \Omega(\frac{1}{Nd})$ instead of $c = \Omega(\frac{1}{N})$). We discuss that in appendix D.

Acknowledgments. I thank Richard Cleve and John Watrous for the discussion that lead me to discovering the main idea of the paper.

References

[1] D. Aharonov. Quantum computation - a review. *Annual Review of Computational Physics*, World Scientific, volume VI. Also quant-ph/9812037.

[2] A. Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64:750-767, 2002. Also quant-ph/0002066.

[3] A. Ambainis. Quantum search algorithms (a survey). *SIGACT News*, 35(2):22-35, 2004. Also quant-ph/0504012.

[4] A. Ambainis. Quantum walk algorithm for element distinctness. *Proceedings of FOCS'04*, pp. 22-31. Also quant-ph/0311001.

[5] A. Ambainis. Quantum search with variable times, quant-ph/0609188.

[6] N. Bshouty, R. Cleve, W. Eberly. Size-depth tradeoffs for algebraic formulas. *SIAM J. Comput.* 24(4): 682-705 (1995)

[7] H. Barnum, M. Saks. A lower bound on the quantum query complexity of read-once functions. *Journal of Computer and System Sciences*, 69(2): 244-258 (2004). Also quant-ph/0201007.

[8] H. Buhrman, R. Cleve, A. Wigderson. Quantum vs. Classical Communication and Computation. *Proceedings of STOC'98*, pp. 63-68. Also quant-ph/9802040.

[9] H. Buhrman, R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288:21-43, 2002.

[10] A. Childs, R. Cleve, S. Jordan, D. Yeung. Discrete-query quantum algorithm for NAND trees, quant-ph/0702160.

[11] A. Childs, B. Reichardt, R. Spalek, S. Zhang. Every NAND formula on N variables can be evaluated in time $O(N^{1/2+\epsilon})$, quant-ph/0703015.

[12] R. Cleve, A. Ekert, C. Machiavello, M. Mosca. On Quantum Algorithms. *Complexity*, 4:33-42, 1998.

[13] E. Farhi, S. Gutmann, An analog analogue of a digital quantum computation. *Physical Review A*, 57:2403, 1997, quant-ph/9612026.

[14] E. Farhi, J. Goldstone, S. Gutmann. A Quantum Algorithm for the Hamiltonian NAND Tree, quant-ph/0702144.

[15] L. Grover. A fast quantum mechanical algorithm for database search. *Proceedings of STOC'96*, pp. 212-219. Also quant-ph/9605043.

[16] P. Høyer, M. Mosca, R. de Wolf. Quantum Search on Bounded-Error Inputs. *Proceedings of ICALP'03*, pp. 291-299. Also quant-ph/0304052.

[17] S. Laplante, T. Lee, M. Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15(2): 163-196, 2006. Also quant-ph/0501057.

[18] C. Mochon. Hamiltonian oracles, quant-ph/0602032.

[19] M. Saks, A. Wigderson. Probabilistic Boolean decision trees and the complexity of evaluating game trees, *Proceedings of FOCS'86*, pp. 29-38.

[20] M. Santha. On the Monte Carlo Boolean Decision Tree Complexity of Read-Once Formulae. *Random Structures and Algorithms*, 6(1): 75-88, 1995.

[21] M. Snir. Lower bounds on probabilistic linear decision trees. *Theoretical Computer Science*, 38: 69-82, 1985.

[22] M. Szegedy. Quantum speed-up of Markov chain based algorithms. *Proceedings of FOCS 2004*, pp. 32-41.

Appendix

A Lemmas on the structure of minimal certificates

Proof: [of Lemma 1] If w belongs to C, then C contains an extended minimal certificate for $T_w = 0$ which, by the argument before the statement of lemma 1 in section 5 (with w instead of v) must contain an extended minimal certificate for one of $T_{y_1} = 0$ and $T_{y_2} = 0$ and an extended minimal certificate for one of $T_{y_3} = 0$ and $T_{y_4} = 0$. From the construction of extended minimal certificates, a certificate for $T_w = 0$ contains y_i if and only if it contains a certificate for $T_{y_i} = 0$. ∎

Proof: [of Lemma 2] Since H_v has non-zero entries only in the places corresponding to the edges of T, the amplitude of each basis state $|u\rangle$ in $H_v|\psi_0\rangle$ is just the sum of amplitudes of the neighbors of u in $|\psi_0\rangle$, multiplied by the appropriate factors. We need to show that, for every u, this sum is 0.

For vertices u in the subtree T_v, if u is at an even depth $2l$, then its neighbors are at an odd depth ($2l-1$ or $2l+1$) and their amplitudes are 0. If u is at an odd depth $2l+1$, let p be the parent of u and let y_1, y_2 be the two children of u. If p does not belong to C, than none of p's descendants belongs to C as well, including y_1 and y_2. Then, all the neighbors of u have amplitude 0 in $|\psi_0\rangle$. If $p \in C$, then by Lemma 1, exactly one of y_1 and y_2 belongs to C. Assume that $y_1 \in C$. Let α be the amplitude of p. By equation (1), the amplitude of y_1 is $-\frac{\sqrt[4]{m_p}}{\sqrt[4]{4m_{y_1}}}\alpha$. The amplitude of u in $H_v|\psi\rangle$ is

$$H_{up}\alpha - H_{uy_1}\frac{\sqrt[4]{m_p}}{\sqrt[4]{4m_{y_1}}}\alpha = \sqrt[4]{\frac{m_p}{2m_u}}\alpha - \sqrt[4]{\frac{2m_{y_1}}{m_u}}\frac{\sqrt[4]{m_p}}{\sqrt[4]{4m_{y_1}}}\alpha = 0.$$

∎

Proof: [of Lemma 3]

Part (a). By induction on the depth of T_v. If C_v is a leaf, then $m_v = 1$ and $\|\psi_{C_v}\| \leq 2\sqrt{1}-1$. For the inductive case, decompose $C_v = \{v\} \cup C_{y_1} \cup C_{y_2}$. By eq. (1) and the inductive assumption, we have

$$\|\psi_{C_v}\|^2 \leq 1 + \frac{\sqrt{m_v}}{2\sqrt{m_{y_1}}}(2\sqrt{m_{y_1}}-1) + \frac{\sqrt{m_v}}{2\sqrt{m_{y_2}}}(2\sqrt{m_{y_2}}-1)$$

$$\leq 2\sqrt{m_v} - \sqrt{m_v}\left(\frac{1}{2\sqrt{m_{y_1}}} + \frac{1}{2\sqrt{m_{y_2}}}\right) + 1 = 2\sqrt{m_v} - 1,$$

with the first inequality following from the inductive assumption, the second inequality following by rearranging terms and the third following from $m_{y_1} = m_{y_2} = \frac{m_v}{2}$ (since the tree is balanced).

Part (b). By induction on the depth of T_v. If C_v is a leaf, then $m_v = d_v = 1$ and $\|\psi_{C_v}\| = 1 \leq 2\sqrt{1}$.

For the inductive case, decompose $C_v = \{v\} \cup C_{y_1} \cup C_{y_2}$. By eq. (1) and the inductive assumption, we have

$$\|\psi_{C_v}\|^2 = 1 + \frac{\sqrt{m_v}}{2\sqrt{m_{y_1}}}\|\psi_{C_{y_1}}\|^2 + \frac{\sqrt{m_v}}{2\sqrt{m_{y_2}}}\|\psi_{C_{y_2}}\|^2$$

$$\leq 1 + \frac{\sqrt{m_v}}{2\sqrt{m_{y_1}}}2\sqrt{m_{y_1}d_{y_1}} + \frac{\sqrt{m_v}}{2\sqrt{m_{y_2}}}2\sqrt{m_{y_2}d_{y_2}} = 1 + \sqrt{m_v d_{y_1}} + \sqrt{m_v d_{y_2}}$$

$$\leq 1 + 2\sqrt{m_v(d_v-1)} \leq \frac{m_v}{d_v} + 2\sqrt{m_v(d_v-1)} \leq 2\sqrt{m_v d_v},$$

with the first inequality following from the inductive assumption, the second inequality following from $d_{y_1} \leq d_v - 1$, $d_{y_2} \leq d_v - 1$, the third inequality following from $d_v \leq m_v$ (the depth of a tree is always at most the number of leaves) and the fourth inequality following from $\sqrt{A} - \sqrt{A-1} = \frac{1}{\sqrt{A}+\sqrt{A-1}} \geq \frac{1}{2\sqrt{A}}$. ∎

Proof: [of Lemma 6] To prove $U_1|\psi_{C_1,C_2}\rangle = |\psi_{C_1,C_2}\rangle$, we need to show $H|\psi_{C_1,C_2}\rangle = 0$. Consider the amplitude of $|u\rangle$ in $H|\psi_{C_1,C_2}\rangle$. If u belongs to T_{v_1} or T_{v_2}, its amplitude in $H|\psi_{C_1,C_2}\rangle$ are 0 by Lemma 2. If $u = v$, its amplitude in $H|\psi_{C_1,C_2}\rangle$ is

$$H_{uv_1}\sqrt[4]{m_{v_1}} - H_{uv_2}\sqrt[4]{m_{v_2}} = \frac{\sqrt[4]{m_u}}{\sqrt[4]{2m_{v_1}}}\sqrt[4]{m_{v_1}} - \frac{\sqrt[4]{m_u}}{\sqrt[4]{2m_{v_2}}}\sqrt[4]{m_{v_2}} = 0.$$

If u is outside T_v, then it has no neighbors in T_v, except for possibly v itself. That means that all of u's neighbors have amplitude 0 in $|\psi_{C_1,C_2}\rangle$ and the amplitude of $|u\rangle$ in $H|\psi_{C_1,C_2}\rangle$ is 0.

For U_2, we have $U_2|\psi_{C_1,C_2}\rangle = |\psi_{C_1,C_2}\rangle$ because the only leaves v with $|v\rangle$ having a non-zero amplitude in $|\psi_{C_1,C_2}\rangle$ are those for which the corresponding variable x_i belongs to C_1 or C_2 and all the variables x_i in a certificate C_j have $x_i = 0$ which means that $U_2|v\rangle = |v\rangle$. ∎

Proof: [of Lemma 11] We prove the lemma for the general case.

The proof is by induction. For the base case, let u be a vertex of depth 1 (i.e. at a distance 1 from a leaf). Then, u is connected to a leaf w. Since w is a leaf, u is the only neighbor of w. This means that the amplitude of w in $H|\psi\rangle$ is equal to $H_{uw}\alpha_u$. Since $H|\psi\rangle = 0$ and $H_{uw} \neq 0$, it must be the case that $\alpha_u = 0$.

For the inductive case, assume that $\alpha_u = 0$ for vertices u of depth $2i+1$, for $i \in \{0, \ldots, l-1\}$. Let u be a vertex at the depth $2l+1$. Let c be one of the two children of u. Let v_1 and v_2 be the children of c. Then, the amplitude of c in $H|\psi\rangle$ is equal to $H_{uc}\alpha_u + H_{cv_1}\alpha_{v_1} + H_{cv_2}\alpha_{v_2}$ and it must be equal to 0. By the inductive assumption, $\alpha_{v_1} = \alpha_{v_2} = 0$. Therefore, $\alpha_u = 0$.

The proof for the vertices in the tail is similar, starting with the vertex that is adjacent to the end of the tail and proceeding inductively towards the root. ∎

B Proof of Lemma 7

Similar statements have been proven before (e.g. [22]) but none of them has the exact form that we need. Therefore, we include the proof for completeness.

Let $|\psi\rangle$ be an eigenvector of U_2U_1 with an eigenvalue λ. We consider the following possibilities:

1. $U_1|\psi\rangle = |\psi\rangle$. Then, $|\psi\rangle$ is an eigenvector of U_2U_1 if and only if it is an eigenvector of U_2. We cannot have $U_2|\psi\rangle = |\psi\rangle$ because, by the conditions of the lemma, $\|P_{(S_2)^\perp}|\psi\rangle\| > 0$. Since all eigenvalues of U_2 are ± 1, this means that $U_2|\psi\rangle = -|\psi\rangle$ and $U_2U_1|\psi\rangle = -|\psi\rangle$.
2. $U_1|\psi\rangle = -|\psi\rangle$. Again, $|\psi\rangle$ is an eigenvector of U_2U_1 if and only if it is an eigenvector of U_2. We cannot have $U_2|\psi\rangle = -|\psi\rangle$ because, then $|\psi\rangle$ would belong to $S_1 \cap S_2$ and the lemma assumes that $S_1 \cap S_2 = \{\overrightarrow{0}\}$. Therefore, $U_2|\psi\rangle = |\psi\rangle$ and $U_2U_1|\psi\rangle = -|\psi\rangle$.
3. $U_1|\psi\rangle \neq |\psi\rangle$ and $U_1|\psi\rangle \neq -|\psi\rangle$.
   Then, $|\psi\rangle$ is not an eigenvector of U_1. This means that $|\psi\rangle$ and $|\psi'\rangle = U_1|\psi\rangle$ span a two dimensional subspace which we denote $\mathcal{H}_2$. Since $U_1^2 = I$, we also have $|\psi\rangle = U_1|\psi'\rangle$. This means that U_1 maps $\mathcal{H}_2$ to itself. U_2 also maps $\mathcal{H}_2$ to itself, because it maps $|\psi'\rangle$ to $U_2|\psi'\rangle = U_2U_1|\psi'\rangle = \lambda|\psi\rangle$ and, since $U_2^2 = I$, this means that $U_2|\psi\rangle = \lambda^{-1}|\psi'\rangle$.
   Therefore, U_2 and U_1 both map $\mathcal{H}_2$ to itself. Let $|\psi_{i1}\rangle$, $|\psi_{i2}\rangle$ be the eigenvectors of U_i in $\mathcal{H}_2$. One of $|\psi_{11}\rangle$, $|\psi_{12}\rangle$ must have an eigenvalue that is +1 and the other must have an eigenvalue -1. (Otherwise, all of $\mathcal{H}_2$, including $|\psi\rangle$, would be eigenvectors of U_1 with the same eigenvalue and then we would have one of the first two cases.) Similarly, one of $|\psi_{21}\rangle$, $|\psi_{22}\rangle$ must have an eigenvalue +1 and the other must have an eigenvalue -1.

For simplicity, assume that $|\psi_{11}\rangle$ and $|\psi_{21}\rangle$ are the eigenvectors with eigenvalue 1. Then, U_2U_1 is a composition of reflections w.r.t. $|\psi_{11}\rangle$ and $|\psi_{21}\rangle$. By the analysis in [1], the eigenvalues of U_2U_1 on $\mathcal{H}_2$ are $e^{\pm i\beta}$, where β is the angle between $|\psi_{11}\rangle$ and $|\psi_{21}\rangle$. We have

$$\|P_{(S_2)^\perp}|\psi_{11}\rangle\|^2 = |\langle\psi_{11}|\psi_{22}\rangle|^2 = \sin^2\beta.$$

By the conditions of the lemma, we have $\sin^2\beta \geq \epsilon$ which implies $\beta \in [\sqrt{\epsilon}, \frac{\pi}{2}]$.

C Evaluating balanced trees: proof of Lemma 10

Since a_v and b_v only depend on k, we will denote them a_k and b_k. We first state some simple bounds on a_k and b_k.

Claim 1 *(a)* $a_k \leq 1.1 \cdot 2^{k+1}$*;*

(b) $b_k \geq 0.9 \cdot \frac{K}{2^k}$*;*

(c) $a_k \leq 0.13 b_k$*;*

Proof: The first two parts follow from $2\frac{2^{2k}}{K} \leq 2\frac{N}{20N} = 0.1$. The third part follows by

$$a_k \leq 1.1 \cdot 2^{k+1} \leq 0.11\frac{K}{2^k} \leq 0.13 b_k,$$

with the second inequality using $K \geq 20N \geq 20 \cdot 2^{2k}$ and the third inequality using part (b). ∎

The proof of Lemma 10 is by an induction on k. The basis case is $k = 0$. Then, the tree consists of the vertex v only. The only possible states $|\psi\rangle$ are multiples of $|v\rangle$. v is also the only leaf, carrying a variable x_1 and $T_0(x_1) = x_1$. If $x_1 = 0$, then $S_{x,1}$ is empty, meaning that $L(\psi) = 1$. If $x_1 = 1$, then $S_{x,1}$ consists of all multiples of $|v\rangle$, meaning that $P_{S_{x,1}}|\psi\rangle = |\psi\rangle$ and $L(\psi) = -(K-1)$. In both cases, the lemma is true.

For the inductive case, let z_1 and z_2 be the children of v, y_1 and y_2 be the children of z_1 and y_3, y_4 be the children of z_2. We can decompose the state $|\psi\rangle$ as

$$|\psi\rangle = \alpha_v|v\rangle + |\psi_1\rangle + |\psi_2\rangle$$

where $|\psi_i\rangle \in S_{z_i,0}$. We claim

Claim 2 *For every $i \in \{1,2\}$, there exists $|\psi_i'\rangle \in S'_{z_i,0}$ such that, for the state $|\psi_i''\rangle = |\psi_i\rangle + |\psi_i'\rangle$, we have*

1. *If $T_{z_1}(x_1, \ldots, x_N) = 0$, then*

$$\|\psi_i''\| - \|P_{x,1}|\psi_i''\rangle\| \leq \frac{-b_{k-1}}{2}|\alpha_v|^2. \tag{2}$$

2. *If $T_{z_1}(x_1, \ldots, x_N) = 1$, then*

$$\|\psi_i''\| - \|P_{x,1}|\psi_i''\rangle\| \leq \frac{a_k - 1}{2}|\alpha_v|^2. \tag{3}$$

Proof: For typographical convenience, let $i = 1$. For the first part, $T_{z_1} = 0$ if and only if $T_{y_1} = T_{y_2} = 1$.

By Lemma 11, the amplitude of $|z_i\rangle$ in $|\psi_1\rangle$ is 0. Therefore, we can decompose

$$|\psi_1\rangle = |\varphi_1\rangle + |\varphi_2\rangle,$$

with $|\varphi_i\rangle \in S_{y_i,0}$. By the inductive assumption, there exist states $|\varphi_1'\rangle, |\varphi_2'\rangle \in S'_{y_i,0}$ such that, for the states $|\varphi_i''\rangle = |\varphi_i\rangle + |\varphi_i'\rangle$, we have

$$\||\varphi_i''\rangle\|^2 - \|P_{S_{x,1}}|\varphi_i''\rangle\|^2 \leq -b_{k-1}|\alpha_{y_i}|^2, \tag{4}$$

with α_{y_i} being the amplitude of y_i in $|\varphi_i\rangle$. We define $|\psi_1'\rangle = |\varphi_1'\rangle + |\varphi_2'\rangle$ and $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$. By summing equations (4) for $i = 1$ and $i = 2$, we get

$$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq -b_{k-1}\sum_{i=1}^{2}|\alpha_{y_i}|^2. \tag{5}$$

Because of $\alpha_{y_1} + \alpha_{y_2} + \alpha_v = 0$, we have $|\alpha_{y_1}|^2 + |\alpha_{y_2}|^2 \geq \frac{|\alpha_v|^2}{2}$. Therefore, (5) implies

$$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq -\frac{b_{k-1}}{2}|\alpha_v|^2.$$

For the second part, we consider two cases:

1. $y_1 = y_2 = 0$.
   Let C be a minimal 0-certificate for $r = 1$ and $|\psi_C\rangle$ be the state corresponding to this certificate. Let $|\psi_1\rangle = \sum_u \alpha_u|u\rangle$. We define

   $$|\phi_1\rangle = \frac{\alpha_{y_1} - \alpha_{y_2}}{2}|\psi_C\rangle.$$

   Let $|\psi_1\rangle + |\phi_1\rangle = \sum_v \beta_v|v\rangle$. Then, $\beta_{y_1} = \beta_{y_2} = \frac{\alpha_{y_1}+\alpha_{y_2}}{2}$. Because of $\alpha_{y_1} + \alpha_{y_2} + \alpha_v = 0$, we have $\beta_{y_1} = \beta_{y_2} = -\frac{\alpha_v}{2}$.
   We decompose

   $$|\psi_1\rangle + |\phi_1\rangle = |\varphi_1\rangle + |\varphi_2\rangle,$$

   with $|\varphi_i\rangle$ being a superposition over T_{y_i}. By the inductive assumption, there exist states $|\varphi_1'\rangle, |\varphi_2'\rangle \in S_{H,0}$ such that, for the states $|\varphi_i''\rangle = |\varphi_i\rangle + |\varphi_i'\rangle$, we have

   $$\||\varphi_i''\rangle\|^2 - \|P_{S_{x,1}}|\varphi_i''\rangle\|^2 \leq a_{k-1}|\alpha_{y_i}|^2. \tag{6}$$

   We define $|\psi_1'\rangle = |\phi_1\rangle + |\varphi_1'\rangle + |\varphi_2'\rangle$ and $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$. Summing up eq. (6) for $i = 1, 2$ gives

   $$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq a_{k-1}(|\beta_{y_1}|^2 + |\beta_{y_2}|^2) = \frac{a_{k-1}}{2}|\alpha_v|^2.$$

   The claim now follows from $a_{k-1} \leq a_k - 1$ which is easy to prove.

2. one of y_1, y_2 is 0 and the other is 1.
   For typographical convenience, assume that $y_1 = 0$, $y_2 = 1$. Once again, we decompose

   $$|\psi_1\rangle = |\varphi_1\rangle + |\varphi_2\rangle,$$

   with $|\varphi_i\rangle$ being a superposition over T_{y_i}. By the inductive assumption, there exist states $|\varphi_1'\rangle, |\varphi_2'\rangle \in S_{H,0}$ such that, for the states $|\varphi_i''\rangle = |\varphi_i\rangle + |\varphi_i'\rangle$, we have (4) for $i = 1$ and (6) for $i = 2$. We define $|\psi_1'\rangle = |\varphi_1'\rangle + |\varphi_2'\rangle$, $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$ and sum up the equations from the inductive assumption. This gives us

   $$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq a_{k-1}|\alpha_{y_1}|^2 - b_{k-1}|\alpha_{y_2}|^2. \tag{7}$$

   Let $|\alpha_{y_2}| = \delta|\alpha_v|$. Then, because of $\alpha_{y_1} + \alpha_{y_2} + \alpha_v = 0$, we have $|\alpha_{y_1}| \leq (1 + \delta)|\alpha_v|$. We now upper-bound the expression

   $$a_{k-1}|\alpha_{y_1}|^2 - b_{k-1}|\alpha_{y_2}|^2 \leq a_{k-1}(1 + \delta)^2|\alpha_v|^2 - b_{k-1}\delta^2|\alpha_v|^2.$$

Let $f(\delta) = a_{k-1}(1+\delta)^2 - b_{k-1}\delta^2$. Then, $f'(\delta) = 2(1+\delta)a_{k-1} - 2\delta b_{k-1}$. The maximum of $f(\delta)$ is achieved when $f'(\delta) = 0$ which is equivalent to $\delta(a_{k-1} - b_{k-1}) = -a_{k-1}$ and $\delta = \frac{a_{k-1}}{b_{k-1}-a_{k-1}}$. Then,

$$f(\delta) = a_{k-1}\left(\frac{b_{k-1}}{b_{k-1}-a_{k-1}}\right)^2 - b_{k-1}\left(\frac{a_{k-1}}{b_{k-1}-a_{k-1}}\right)^2 = \frac{a_{k-1}b_{k-1}}{b_{k-1}-a_{k-1}}.$$

This means that

$$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq \frac{a_{k-1}b_{k-1}}{b_{k-1}-a_{k-1}}|\alpha_v|^2.$$

To complete the case, it suffices to show

$$2\frac{a_{k-1}b_{k-1}}{b_{k-1}-a_{k-1}} + 1 \leq a_k. \tag{8}$$

We have

$$\begin{aligned} 2\frac{a_{k-1}b_{k-1}}{b_{k-1}-a_{k-1}} &= 2\left(1+\frac{a_{k-1}}{b_{k-1}-a_{k-1}}\right)a_{k-1} \leq 2\left(1+\frac{1.1\cdot 2^k}{0.87 b_k}\right)a_{k-1} \\ &\leq 2\left(1+\frac{1.1\cdot 2^k}{0.87\cdot 0.9\frac{K}{2^{k-1}}}\right)a_{k-1} \leq 2\left(1+1.41\frac{2^{2k-1}}{K}\right)a_{k-1} \\ &\leq 2\left(1+2.82\frac{2^{2k-2}}{K}\right)\left(1+2\frac{2^{2k-2}}{K}\right)(2^k-1) \leq \left(1+2\frac{2^{2k}}{K}\right)(2^{k+1}-2), \end{aligned} \tag{9}$$

with the first inequality following from parts (a) and (c) of Claim 1, the second inequality following from part (b) of Claim 1, the fourth inequality following by writing out a_{k-1} and the last inequality following from $(1+2\delta)(1+2.82\delta) \leq 1+8\delta$ (where $\delta = \frac{2^{2k-2}}{K}$) being true for sufficiently small δ. The equation (8) now follows by adding 1 to both sides of eq. (9).

To deduce lemma 10 from claim 2, we define

$$|\psi'\rangle = |\psi_1'\rangle + |\psi_2'\rangle.$$

Let $|\psi''\rangle = |\psi\rangle + |\psi'\rangle$. Then, we also have

$$|\psi''\rangle = \alpha_v|v\rangle + |\psi_1''\rangle + |\psi_2''\rangle.$$

If $T_r = 0$, then $T_{z_1} = T_{z_2} = 1$. By summing up eq. (3) for $i = 1, 2$ and adding $|\alpha_v|^2$ to both sides, we get

$$\||\psi''\rangle\| - \|P_{x,1}|\psi''\rangle\| \leq a_k|\alpha_v|^2.$$

If $T_r = 1$, then we again have two cases:

1. $T_{z_1} = T_{z_2} = 1$.
   By summing up eq. (2) for $i = 1, 2$ and adding $|\alpha_v|^2$ to both sides, we get
   $$\||\psi''\rangle\| - \|P_{x,1}|\psi''\rangle\| \leq -(b_{k-1}-1)|\alpha_v|^2.$$
   The lemma follows from $b_{k-1} - 1 \geq b_k$ which is easy to prove.
2. one of z_1, z_2 is 0 and the other is 1.
   For typographical convenience, assume that $z_1 = 0$, $z_2 = 1$. In this case, claim 2 gives us
   $$\||\psi''\rangle\| - \|P_{x,1}|\psi''\rangle\| \leq -\frac{b_{k-1}-a_k-1}{2}|\alpha_v|^2.$$
   To complete the proof, we need to show that $\frac{b_{k-1}-a_k-1}{2} \geq b_k$. This follows by substituting the expressions for a_k, b_{k-1} and b_k.

D General case

The counterpart of Lemma 10 is

Lemma 12 *Let $K = 30Nd$. For $v \in T$, define $\delta_v = \frac{5m_v\sqrt{d_v}}{K} + \frac{d_v}{\sqrt{K}}$, where d_v is the depth of the subtree T_v. For any v of even depth and any state $|\psi\rangle \in S_{v,0}$, there exists a state $|\psi'\rangle \in S'_{v,0}$ such that, for $|\psi''\rangle = |\psi\rangle + |\psi'\rangle$, we have*

1. *If $T_v(x_1, \ldots, x_N) = 0$, then*

$$L_v(\psi) \leq a_v, \quad a_v = 2(1+\delta_v)\sqrt{d_v m_v}.$$

2. *If $T_k(x_1, \ldots, x_N) = 1$, then*

$$L_v(\psi) \leq -b_v, \quad b_v = (1-\delta_v)\frac{K}{\sqrt{m_v}}.$$

Observe that, because of $m_v \leq N$ and $d_v \leq d \leq N$, we always have

$$\delta_v \leq \frac{5N\sqrt{d}}{30Nd} + \frac{d}{30\sqrt{dN}} \leq \frac{5}{30} + \frac{1}{30} = \frac{1}{5}.$$

Proof: By induction on the depth l of the subtree T_v. The basis case is $l = 0$. Then, the tree consists of v only. The only possible states $|\psi\rangle$ are multiples of $|v\rangle$. The root is also the only leaf, carrying a variable x_i and $T_v(x_i) = x_i$. If $x_i = 0$, then $S_{x,1}$ is empty, meaning that $L_v(\psi) = 1$. If $x_i = 1$, then $S_{x,1}$ consists of all multiples of $|r\rangle$, meaning that $P_{S_{x,1}}|\psi\rangle = |\psi\rangle$ and $L_v(\psi) = -(K-1)$. In both cases, the lemma is true.

For the inductive case, let z_1 and z_2 be the children of v, y_1 and y_2 be the children of z_1 and y_3, y_4 be the children of z_2. We can decompose the state $|\psi\rangle$ as

$$|\psi\rangle = \alpha_v|v\rangle + |\psi_1\rangle + |\psi_2\rangle \tag{10}$$

where $|\psi_i\rangle$ is a superposition over $|u\rangle, u \in T_{z_i}$. We claim

Claim 3 *For every $i \in \{1,2\}$, there exists $|\psi'_i\rangle \in S'_{z_i,0}$, such that, for the state $|\psi''_i\rangle = |\psi_i\rangle + |\psi'_i\rangle$, we have*

1. *If $T_{z_1}(x_1, \ldots, x_N) = 0$, then*

$$\||\psi''_i\|^2 - \|P_{x,1}|\psi''_i\rangle\|^2 \leq \frac{-b_{y_i}}{\sqrt{2}}|H_{vy_i}\alpha_v|^2. \tag{11}$$

2. *If $T_{z_1}(x_1, \ldots, x_N) = 1$, then*

$$\||\psi''_i\|^2 - \|P_{x,1}|\psi''_i\rangle\|^2 \leq \frac{a_{y_i}}{\sqrt{2}}|H_{vy_i}\alpha_v|^2. \tag{12}$$

Proof: For typographical convenience, let $i = 1$. By Lemma 11, the amplitude of z_1 in $|\psi_1\rangle$ is 0. Therefore, we can decompose

$$|\psi_1\rangle = |\varphi_1\rangle + |\varphi_2\rangle,$$

with $|\varphi_i\rangle$ being a superposition over T_{y_i}.

For the first part, $T_{z_1} = 0$ if and only if $T_{y_1} = T_{y_2} = 1$. By the inductive assumption, there exist states $|\varphi'_1\rangle \in S'_{y_1,0}$, $|\varphi'_2\rangle \in\in S'_{y_2,0}$ such that, for the states $|\varphi''_i\rangle = |\varphi_i\rangle + |\varphi'_i\rangle$, we have

$$\||\varphi''_i\|^2 - \|P_{S_{x,1}}|\varphi''_i\rangle\|^2 \leq -b_{y_i}|\alpha_{y_i}|^2, \tag{13}$$

with α_{y_i} being the amplitude of y_i in $|\varphi_i\rangle$ (which is the same as its amplitude in $|\psi_1\rangle$). We define $|\psi_1'\rangle = |\varphi_1'\rangle + |\varphi_2'\rangle$ and $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$. By summing equations (13) for $i = 1$ and $i = 2$, we get

$$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq -\sum_{i=1}^{2} b_{y_i}|\alpha_{y_i}|^2. \tag{14}$$

We would like to upperbound the right-hand side of this equation. From $H|\psi\rangle = 0$, we have

$$H_{y_1 z_1}\alpha_{y_1} + H_{y_2 z_1}\alpha_{y_2} + H_{v z_1}\alpha_v = 0. \tag{15}$$

Define $x_i = -\frac{H_{y_i z_1}\alpha_{y_i}}{H_{v z_1}\alpha_v}$. Then, by dividing both sides of (15) by $-H_{v z_1}\alpha_v$, we have $x_1 + x_2 = 1$. By expressing α_{y_i} in terms of x_i, we get

$$-\sum_{i=1}^{2} b_{y_i}|\alpha_{y_i}|^2 = -|H_{v z_1}\alpha_v|^2 \sum_{i=1}^{2} \frac{b_{y_i}}{H_{y_i z_1}^2}|x_i|^2$$

$$\leq -|H_{v z_1}\alpha_v|^2(1-\delta_{z_1}) \sum_{i=1}^{2} \frac{K\sqrt{m_{z_1}}}{\sqrt{2}m_{y_i}}|x_i|^2,$$

where the last inequality follows by writing out b_{y_i} and $H_{y_i z_1}$ and applying $\delta_{y_i} \leq \delta_{z_1}$ (which is true because both the size and the depth of T_{y_i} are less than the size and the depth of T_{z_1}). To complete the proof, it suffices to show that

$$\frac{|x_1|^2}{m_{y_1}} + \frac{|x_2|^2}{m_{y_2}} \geq \frac{1}{m_{z_1}}, \tag{16}$$

subject to the constraint $x_1 + x_2 = 1$. The left hand side of (16) is minimized when x_1 and x_2 are both real. (Otherwise, one can replace x_1 and x_2 by $x_1' = \frac{|x_1|}{|x_1|+|x_2|}$ and $x_2' = \frac{|x_2|}{|x_1|+|x_2|}$ and this does not increase the left hand side.) Therefore, we can find the minimum of the left hand side of (16) by substituting $x_2 = 1 - x_1$ and taking the derivative of the left hand side. That shows that the left hand side is minimized by $x_1 = \frac{m_{y_1}}{m_{y_1}+m_{y_2}}$, $x_2 = \frac{m_{y_2}}{m_{y_1}+m_{y_2}}$. Then, it is equal to $\frac{1}{m_{y_1}+m_{y_2}} = \frac{1}{m_{z_1}}$.

For the second part, we consider two cases:

1. $y_1 = y_2 = 0$.

   Let C_1, C_2 be extended minimal certificates for $T_{y_1} = 0$ and $T_{y_2} = 0$, respectively and $\gamma = \frac{\alpha_{y_2}-\alpha_{y_1}}{\sqrt[4]{m_{y_1}}+\sqrt[4]{m_{y_2}}}$. We define

   $$|\tilde{\varphi}_1\rangle = |\varphi_1\rangle + \gamma\sqrt[4]{m_{y_1}}|\psi_{C_1}\rangle.$$

   We define $|\tilde{\varphi}_2\rangle$ similarly, with - sign in the front of $\sqrt{m_{y_2}}|\psi_{C_2}\rangle$. We then apply the inductive assumption to the states $|\tilde{\varphi}_i\rangle$, obtaining $|\varphi_1'\rangle, |\varphi_2'\rangle$ such that for $|\varphi_i''\rangle = |\tilde{\varphi}_i\rangle + |\varphi_i'\rangle$, we have

   $$\||\varphi_i''\rangle\|^2 - \|P_{S_{x,1}}|\varphi_i''\rangle\|^2 \leq a_{y_i}|\beta_{y_i}|^2, \tag{17}$$

   where β_{y_i} is the amplitude of y_i in $|\tilde{\varphi}_i\rangle$. We define

   $$|\psi_1'\rangle = |\varphi_1'\rangle + |\varphi_2'\rangle + \gamma|\psi_{C_1,C_2}\rangle$$

   and let $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$. Then, by summing equations (17), we get

   $$\||\psi_1''\rangle\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq a_{y_1}|\beta_{y_1}|^2 + a_{y_2}|\beta_{y_2}|^2. \tag{18}$$

By our choice, we have $\beta_{y_1} = \beta_{y_2}$. Since $H|\psi\rangle = 0$ and $H|\psi'\rangle = 0$, we have $H|\psi''\rangle = 0$. By writing out the amplitude of $|z_1\rangle$ in $H|\psi''\rangle$, we get

$$H_{y_1 z_1}\beta_{y_1} + H_{y_2 z_1}\beta_{y_2} + H_{z_1 v}\alpha_v = 0.$$

Therefore,

$$\beta_{y_1} = \beta_{y_2} = -\frac{H_{z_1 v}\alpha_v}{H_{y_1 z_1} + H_{y_2 z_1}}.$$

By substituting that into (18), we get

$$\|\psi_1''\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq \frac{a_{y_1} + a_{y_2}}{(H_{y_1 z_1} + H_{y_2 z_1})^2}|H_{z_1 v}\alpha_v|^2.$$

We now expand the coefficient of $|H_{z_1 v}\alpha_v|^2$:

$$\frac{a_{y_1} + a_{y_2}}{(H_{y_1 z_1} + H_{y_2 z_1})^2} \leq 2(1+\delta_{z_1})\frac{\sqrt{d_{y_1} m_{y_1}} + \sqrt{d_{y_2} m_{y_2}}}{\left(\frac{\sqrt[4]{2m_{y_1}}}{\sqrt[4]{m_{z_1}}} + \frac{\sqrt[4]{2m_{y_2}}}{\sqrt[4]{m_{z_1}}}\right)^2}$$

$$\leq \frac{1}{\sqrt{2}}2(1+\delta_{z_1})\sqrt{d_{z_1} m_{z_1}}\frac{\sqrt{m_{y_1}} + \sqrt{m_{y_2}}}{(\sqrt[4]{m_{y_1}} + \sqrt[4]{m_{y_2}})^2}$$

$$\leq \frac{1}{\sqrt{2}}2(1+\delta_{z_1})\sqrt{d_{z_1} m_{z_1}}.$$

2. one of y_1, y_2 is 0 and the other is 1.

   For typographical convenience, assume that $y_1 = 0$, $y_2 = 1$. By the inductive assumption, there exist states $|\varphi_1'\rangle, |\varphi_2'\rangle \in S_{H,0}$ such that, for the states $|\varphi_i''\rangle = |\varphi_i\rangle + |\varphi_i'\rangle$, we have

$$\|\varphi_i''\|^2 - \|P_{S_{x,1}}|\varphi_i''\rangle\|^2 \leq c|\alpha_{y_i}|^2, \tag{19}$$

   with $c = a_{y_1}$ for $i = 1$ and $c = -b_{y_2}$ for $i = 2$. We define $|\psi_1'\rangle = |\varphi_1'\rangle + |\varphi_2'\rangle$ and $|\psi_1''\rangle = |\psi_1\rangle + |\psi_1'\rangle$. By summing the equations (19), we get

$$\|\psi_1''\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq a_{y_1}|\alpha_{y_1}|^2 - b_{y_2}|\alpha_{y_2}|^2. \tag{20}$$

   We switch to variables $x_i = -\frac{H_{y_i z_1}\alpha_{y_i}}{H_{v z_1}\alpha_v}$, obtaining

$$\|\psi_1''\|^2 - \|P_{S_{x,1}}|\psi_1''\rangle\|^2 \leq |H_{v z_1}\alpha_v|^2\left(\frac{a_{y_1}}{H_{y_1 z_1}^2}|x_1|^2 - \frac{b_{y_2}}{H_{y_2 z_1}^2}|x_2|^2\right).$$

   Substituting the expressions for $a_{y_1}, b_{y_2}, H_{y_1 z_1}, H_{y_2 z_1}$ gives

$$\frac{a_{y_1}}{H_{y_1 z_1}^2}|x_1|^2 - \frac{b_{y_2}}{H_{y_2 z_1}^2}|x_2|^2$$

$$= \frac{1}{\sqrt{2}}\left((1+\delta_{y_1})2\sqrt{d_{y_1} m_{z_1}}|x_1|^2 - (1-\delta_{y_2})\frac{K\sqrt{m_{z_1}}}{m_{y_2}}|x_2|^2\right)$$

$$\leq \frac{1}{\sqrt{2}}2(1+\delta_{y_1})\sqrt{d_{y_1} m_{z_1}}\left(|x_1|^2 - \frac{K}{3\sqrt{d_{y_1}} m_{y_2}}|x_2|^2\right), \tag{21}$$

   with the second inequality following by rearranging terms and applying $1 - \delta_{y_2} \geq \frac{4}{5} \geq \frac{4}{6}(1+\delta_{y_1})$ which is true because of $\delta_{y_1} \leq \frac{1}{5}$ and $\delta_{y_2} \leq \frac{1}{5}$.

Let $\delta = |x_2|$. Then $|x_1| \le 1 + \delta$ and the right hand side of (21) is at most

$$\frac{1}{\sqrt{2}} 2(1+\delta_{y_1})\sqrt{d_{y_1} m_{z_1}} \left((1+\delta)^2 - \frac{K}{3\sqrt{d_{y_1}} m_{y_2}} \delta^2 \right).$$

Taking the derivative w.r.t. δ shows that the expression in the brackets is maximized by

$$\delta = \frac{3\sqrt{d_{y_1}} m_{y_2}}{K - 3\sqrt{d_{y_1}} m_{y_2}},$$

in which case it is equal to $\frac{K}{K-3\sqrt{d_{y_1}} m_{y_2}}$. To complete the proof, we observe that

$$(1+\delta_{y_1}) \frac{K}{K - 3\sqrt{d_{y_1}} m_{y_2}} = 1 + \delta_{y_1} + \frac{(1+\delta_{y_1}) 3\sqrt{d_{y_1}} m_{y_2}}{K - 3\sqrt{d_{y_1}} m_{y_2}}$$

$$\le 1 + \delta_{y_1} + 4\frac{\sqrt{d_{y_1}} m_{y_2}}{K} < 1 + \delta_{y_1} + 4\frac{\sqrt{d_{z_1}} m_{y_2}}{K} < 1 + \delta_{z_1},$$

with the first inequality following from $\delta_{y_1} < \frac{1}{5}$ and $K - 3\sqrt{d_{y_2}} m_{y_2} \ge K - 3N\sqrt{d} \le \frac{27}{30} K$.

To deduce lemma 12 from claim 3, we define

$$|\psi'\rangle = |\psi'_1\rangle + |\psi'_2\rangle.$$

Because of equation (10), we can also express $|\psi''\rangle = |\psi\rangle + |\psi'\rangle$ as

$$|\psi''\rangle = \alpha_v |v\rangle + |\psi''_1\rangle + |\psi''_2\rangle.$$

If $T_r = 0$, then $T_{z_1} = T_{z_2} = 1$. By summing up eq. (12) for $i = 1, 2$ and adding $|\alpha_v|^2$ to both sides, we get

$$\||\psi''\|^2 - \|P_{x,1}|\psi''\rangle\|^2 \le \left(\frac{a_{z_1}}{\sqrt{2}} H_{vz_1}^2 + \frac{a_{z_2}}{\sqrt{2}} H_{vz_2}^2 + 1 \right) |\alpha_v|^2. \tag{22}$$

By expanding a_{z_i} and H_{vz_i}, we get

$$\frac{a_{z_1}}{\sqrt{2}} H_{vz_1}^2 + \frac{a_{z_2}}{\sqrt{2}} H_{vz_2}^2 + 1 \le (1+\delta_v) \left(\frac{m_{z_1}\sqrt{d_{z_1}}}{\sqrt{m_v}} + \frac{m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + 1 \right).$$

Since $m_v = m_{z_1} + m_{z_2}$, we have $\frac{m_{z_1}}{\sqrt{m_v}} + \frac{m_{z_2}}{\sqrt{m_v}} = \frac{m_v}{\sqrt{m_v}} = \sqrt{m_v}$. Together with $d_v = \max(d_{z_1}, d_{z_2}) + 1$, this implies

$$\frac{m_{z_1}\sqrt{d_{z_1}}}{\sqrt{m_v}} + \frac{m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + 1 \le \sqrt{m_v(d_v - 1)} + 1$$

$$\le \sqrt{m_v(d_v-1)} + \frac{\sqrt{m_v}}{\sqrt{d_v}} = \sqrt{m_v} \frac{2\sqrt{d_v(d_v-1)} + 1}{\sqrt{d_v}}$$

$$\le \sqrt{m_v} \frac{2d_v}{\sqrt{d_v}} = 2\sqrt{m_v d_v},$$

with the second inequality following from $d_v \le m_v$ (the depth of any tree is at most the number of leaves in it) and the last inequality following from $\sqrt{d_v(d_v-1)} \le d_v - \frac{1}{2}$.

If $T_r = 1$, then we again have two cases:

1. $T_{z_1} = T_{z_2} = 1$.
   By summing up eq. (11) for $i = 1, 2$ and adding $|\alpha_v|^2$ to both sides, we get

$$\|\psi''\|^2 - \|P_{x,1}|\psi''\rangle\|^2 \leq -\left(\frac{b_{z_1}}{\sqrt{2}}H_{vz_1}^2 + \frac{b_{z_2}}{\sqrt{2}}H_{vz_2}^2 - 1\right)|\alpha_v|^2.$$

   By expanding b_{z_i} and H_{vz_i} and simplifying, we get

$$\frac{b_{z_1}}{\sqrt{2}}H_{vz_1}^2 + \frac{b_{z_2}}{\sqrt{2}}H_{vz_2}^2 - 1 \geq (1-\delta_{z_1})\frac{K}{\sqrt{m_v}} + (1-\delta_{z_2})\frac{K}{\sqrt{m_v}} - 1$$

$$\geq 2(1-\delta_v)\frac{K}{\sqrt{m_v}} - 1 > (1-\delta_v)\frac{K}{\sqrt{m_v}}.$$

2. one of z_1, z_2 is 0 and the other is 1.
   For typographical convenience, assume that $z_1 = 0$, $z_2 = 1$. In this case, claim 3 gives us

$$\|\psi''\|^2 - \|P_{x,1}|\psi''\rangle\|^2 \leq \left(-\frac{b_{z_1}}{\sqrt{2}}H_{vz_1}^2 + \frac{a_{z_2}}{\sqrt{2}}H_{vz_2}^2 + 1\right)|\alpha_r|^2.$$

   We have

$$-\frac{b_{z_1}}{\sqrt{2}}H_{vz_1}^2 + \frac{a_{z_2}}{\sqrt{2}}H_{vz_2}^2 + 1 \leq -(1-\delta_{z_1})\frac{K}{\sqrt{m_v}} + 2(1+\delta_{z_2})\frac{m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + 1.$$

   To prove that this is at most $-b_v$, we need to show that

$$2(1+\delta_{z_2})\frac{m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + 1 \leq (\delta_v - \delta_{z_1})\frac{K}{\sqrt{m_v}}. \tag{23}$$

   This follows from

$$(\delta_v - \delta_{z_1})\frac{K}{\sqrt{m_v}} = \left(\frac{5m_v\sqrt{d_v}}{K} + \frac{d_v}{\sqrt{K}} - \frac{5m_{z_1}\sqrt{d_{z_1}}}{K} - \frac{d_{z_1}}{\sqrt{K}}\right)\frac{K}{\sqrt{m_v}}$$

$$\geq \left(\frac{5m_{z_2}\sqrt{d_v}}{K} + \frac{1}{\sqrt{K}}\right)\frac{K}{\sqrt{m_v}} \geq \frac{5m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + \frac{\sqrt{K}}{\sqrt{m_v}}$$

$$\geq 2(1+\delta_{z_2})\frac{m_{z_2}\sqrt{d_{z_2}}}{\sqrt{m_v}} + 1.$$

   The first equality follows by writing out δ_v and δ_{z_1}, the next inequality follows from $m_v = m_{z_1} + m_{z_2}$ and $d_v > d_{z_1}$ and the last inequality follows from $\delta_{z_2} < \frac{1}{5}$ and $m_v \leq N \leq \frac{K}{30}$.

This completes the proof of Lemma 12. ∎

The general case of Lemma 9 now follows from Lemma 12 in the same way as the balanced case of Lemma 9 followed from Lemma 10.

Distinction between balanced and general case. There are two reasons why our bound for the general case has $O(\sqrt{Nd})$ instead of $O(\sqrt{N})$ for the balanced case:

1. $z_1 = 0$, $z_2 = 1$ case of the proof of Lemma 12 which requires having $\sqrt{d_v m_v}$ instead of $\sqrt{m_v}$. (The rest of the proof of Lemma 12 would still work with $\sqrt{m_v}$.)
2. For the $T = 0$ case, lemma 3 has $O(\sqrt{m_v d_v})$ in the general case and $O(\sqrt{m_v})$ in the balanced case.